

DEVELOPMENT IN L2 WRITING AFTER A SEMESTER OF STUDY IN AN AUSTRALIAN UNIVERSITY

Neomy Storch

*School of Languages and Linguistics
The University of Melbourne*

Abstract

There is a common expectation, particularly amongst international students, that studying in an English-medium university would lead to an improvement in their writing skills (as well as other language skills). However, to date there has been very little research about the impact of studying in an English medium university on the development of learners' writing. This paper reports on study which was part of a larger project. The larger project, using a test-re-test design, investigated whether the reading and writing skills of international students improved after one semester of study in an Australian university. The study reported here analysed the writing test scripts of 20 students whose global writing scores indicated improvement in writing and who had not accessed any language support during the semester. A range of quantitative and qualitative measures were used to analyse the data, including measures of linguistic fluency, accuracy and complexity. The study found that the greatest improvements occurred in how learners structured their writing and developed their ideas. There was also a marked improvement in the formality of learners' language, but less evidence of improvement in language accuracy and complexity. These findings suggest that some aspects of written language may need more explicit language instruction in order to improve.

Keywords: English medium university, writing test scripts, linguistic fluency, linguistic accuracy, linguistic complexity, explicit language instruction.

INTRODUCTION

In recent years there has been a dramatic increase in the number of international students in Australian universities, reflecting a trend observed in many other universities in English speaking countries. One of the assumed advantages of studying at a university of the target language is that of improved second language (L2) skills. Studying in the L2 medium

university offers the learner an opportunity for plentiful exposure to and engagement with the L2. Theories of second language acquisition have long ago identified the importance of rich and authentic input (e.g. Krashen, 1985) and the need to produce appropriate and meaningful language output (Swain, 1985) for L2 learning.

However, studies investigating the relationship between English language tuition in English medium institutions and development of learners' language skills have produced somewhat mixed results. For example, in a small-scale (n=17) study in New Zealand, Read and Hays (2003) found that gains made by their students following one month of an IELTS preparation course were not statistically significant. In a large-scale study (n=476) conducted in the United Kingdom, Green and Weir (2003) found that, on average, students' scores only increased by 0.21 of a band (from an average of score of 5.27 to 5.48) following 3-12 weeks of intensive IELTS preparation and English for Academic Purposes type courses. In contrast, Elder and O'Loughlin (2003) found that 10-12 weeks of intensive English language courses in Australia and New Zealand (n=112) resulted in a significant improvement in English language proficiency, with students on average increasing their IELTS score by half a band. However, the researchers found that improvement was greatest on the Listening subtest and gains were likely to be greater for students with low initial English language proficiency. They also found that improvements were evident in average gains for the entire cohort but that individual performances varied considerably.

Other studies investigated learners' writing development using a range of measures of writing fluency, accuracy, and complexity rather than just relying on global proficiency scores. For example, Shaw and Liu (1998) compared international students' writing before and after they completed a full time English for Academic Purposes (EAP) course in the United Kingdom. The researchers focused mainly on the linguistic features of the learners' writing. They reported that the learners' writing became more formal, employing language associated with written rather than spoken register, but showed no significant changes in terms of linguistic accuracy and complexity.

It should be noted that most of these empirical studies have been confined largely to investigations of the effects of test preparation or intensive English for Academic Purposes type courses, completed prior to university entrance. Very few studies have investigated the impact of studying in an English medium institution on language development. Hinkel's (2003) large scale study compared the texts produced by native speakers and non-native speaker students in US universities. The non-native speakers were considered advanced learners because they had spent at least

four years at Community Colleges and High School in the United States. Hinkel found that, despite their in-country experience, the written texts produced by the non-native speakers were still less linguistically sophisticated than those produced by English native speakers. In a large research project we conducted in an Australian university, results showed that, on average, the learners' reading and writing skills, as measured by global scores, had improved over the course of a semester (Storch & Hill, under review). The improvement was statistically significant. However, not all students showed improvement. It was also interesting to note, that although the University offered a range of English language support, only about a third of the students in our cohort reported that they had accessed this support.

The present study used a subset of the data from the larger project. The study aimed to investigate what aspects of academic writing improved after one semester of study. I focused specifically on students who showed improvement in their writing (based on their global writing scores) and who had reported in their questionnaires that they did not access any academic English language support. The aim of this study was to investigate what aspects of writing showed development after studying in an L2 context. The study investigated L2 writing development in terms of text structure, content, and language. However, I was particularly interested in investigating linguistic aspects of writing, that is, whether the learners' writing became more fluent, accurate and complex as a result of studying in an English medium university.

STUDY DESIGN

The study used a test-re-test design. In the larger project from which data for this study was drawn, 39 participants completed a diagnostic reading and writing test twice: at the beginning and end of their first semester at the university (i.e. after 12 weeks). The diagnostic test, entitled Diagnostic English Language Assessment (DELA) has been developed by the University, and is used to identify students who may benefit from further language development. The test is available free of charge for all incoming international students.

The test has three subtests: listening, reading, and writing. Scores on each subtest are reported on a scale of 1 to 9, with 9 representing an advanced level of proficiency. The writing subtest is assessed on the criteria of Communicative Fluency, Content, and Form (vocabulary and grammatical accuracy). For each of these three criteria the student receives a score of 1 to 9 (using a descriptive scale), and the three scores are then averaged to yield a single, global writing score.

The test results are used to generate recommendations regarding the type of language support, if any, the student is likely to need. A range of support options is available: ESL credit subjects, not for credit workshops and short courses as well as individual consultations. However, neither the test nor its recommended support options are mandatory.

Participants

The data of 20 participants were selected for this study. The participants were chosen because their global writing scores had shown an improvement. Furthermore, these participants had indicated on the questionnaire they completed as part the larger project, that they had not attended an ESL course nor sought help with their English during their semester of study at the University.

All 20 participants (13 female, 7 male) were from Asia, from countries such as China, Indonesia and Vietnam. As all students enrolling at the university must fulfil certain English language requirements, the participants represented a fairly narrow range of proficiency from intermediate to upper-intermediate (e.g., Overall IELTS scores in the range of 6.5-7). Most of the participants were postgraduate students, mainly from the Faculties of Economics/Commerce and Engineering.

Data

The DELA written test (scores and scripts) produced at the beginning of semester (Time 1) and again towards the end of the semester (Time 2) were the main source of data. The same version of the test was used on both occasions. The writing test was double-marked with any discrepancies resolved through discussion. The writing task asked students to write an argumentative essay of at least 300 words on the topic of animal rights. Students were given 5 minutes to read six short excerpts, 5 minutes to plan the essay and 45 minutes to write.

Data Analysis

Since all scripts were selected because they had demonstrated an improved global writing score at Time 2, the analysis focused more closely on criteria scores (Communicative Fluency, Content, Form), noting aspects of communicative fluency (organisation of ideas), content (number and development of ideas), and language use (grammar and lexis). Analysis of language use utilised a range of measures of writing fluency, accuracy, syntactic and lexical complexity. These measures required all essays to be

coded for length (in words), T-units, clauses, and errors, and to consider word choice more closely.

A T-unit is defined by Hunt (1966, p. 735) as “one main clause plus whatever subordinate clauses happen to be attached to or embedded within it.” Written scripts were also coded for clauses, distinguishing between independent and dependent clauses. An independent clause is one that can be used on its own (Richards et al., 1992). In this study a dependent clause was one which contained a finite verb and a subject (Wolf-Quintero et al., 1998). For example, the following T-unit from the data contains 2 clauses (shown separated by a slash) an independent clause and a dependent clause (beginning with *just because*):

People can not become all vegetarians/ just because they have to respect animal rights.

Fluency was measured in terms of number of words. Wolf-Quintero et al., (1998), based on an extensive review of studies, argue that word count is a reliable measure of writing fluency when writing is composed under time constraints. Another reliable measure of fluency is words per T-unit.

In order to assess accuracy, all essays were coded for errors. Errors were also classified, distinguishing between errors in syntax (e.g., errors in word order, missing elements), errors in morphology (e.g., verb tense, subject-verb agreement, use of articles), and errors in lexis (word choice). All errors in spelling and punctuation were ignored. A range of accuracy scores were then calculated: a ratio of error free T-units per T-unit (EFT/T), a ratio of error free clauses per clause (EFC/C), and the total number of errors per total number of words (E/W). The last measure (E/W) addresses the concern that some researchers have raised about using only ratio scores to assess learners' accuracy (Bardovi-Halig & Bofman, 1989); that is, that ratio scores do not distinguish between units (T-units or clauses) which contain multiple errors from those which contain only a single error.

Both syntactic and lexical complexity were measured. Two measures of grammatical complexity were used in this study. The first was the ratio of clauses to T-units (C/T). The other was the proportion of dependent clauses to all clauses (DC/C), (Wolf-Quintero et al., 1998). Both measures are based on the assumption that a move from coordination to subordination and embedding reflects greater syntactic complexity (Foster & Skehan, 1999). To measure lexical complexity, average length of words was calculated by dividing the total number of characters by the total number of words in the text (Cummins et al., 2005)¹. In addition, drawing on research

which has identified some of the features which distinguish academic writing from informal writing (e.g., Biber, 1988; Swales, 1985), the essays were also analysed for type of words and expressions. A closer analysis of the essays identified words and phrases considered colloquial and informal and thus more suitable in spoken rather than written English (Hinkel, 2003). These included the use of contractions (e.g., I'm, it's, etc.), vague words (e.g., people, things), informal words and expressions (e.g., a lot, lots of, kind of, stuff), informal links (e.g., so, by the way), use of first person pronouns (I, me, my), addressing the reader (e.g., you should consider...), and inappropriate use of rhetorical questions (e.g., Who say animals have more rights?) and exclamations (e.g., That example is worse than jail!). The essays were also analysed for the presence of what were considered more formal perhaps and less frequently used words (e.g., phenomenon, genetic, eliminate). Such rare words, used appropriately, may be considered markers of lexical richness (Hinkel, 2003). Figure 1 summarises the measures used to analyse the language used in the essays.

<i>Fluency</i>	<i>Accuracy</i>	<i>Complexity</i>	
		<i>Grammatical</i>	<i>Lexical</i>
Total No. of words (W)	Error free T units per T-unit (EFT/T)	Clauses per T-unit (C/T)	Average word length
No words/ T-unit (W/T)	Error-free clauses per clauses (EFC/C)	Dependent clauses per clause (DC/C)	No. of informal expression
	No. errors/total words (E/W)		No. of rare /formal words

Figure 1: Measures used in analysis of written scripts

In order to check for inter- and intra-rater reliability in coding, and following the advice of Polio (1997), guidelines were formulated stating clearly what constitutes a T-unit, a clause and an error. Then, a random sample of four writing scripts were coded by a second researcher. Intra-rater reliability for T-unit and clause identification was 0.94 and 0.88 respectively. Inter-rater reliability for error counts was 0.82. Discussion between the researchers resolved all disagreements.

RESULTS

Communicative Fluency

The criterion of Communicative Fluency includes reference to how well the writing is structured (in terms of ease of reading) and to text cohesion and coherence. The writing of 16 students showed improvement

on this criterion score, four of which showed a two-band gain from Time 1 to Time 2. In analysing the writing for aspects of Communicative Fluency, I considered the introductions and conclusions and the links established between and within ideas.

All the essays at Time 1 and Time 2 began with some sort of introduction, with the majority (75%) including a brief general statement describing the controversy followed by some sort of a thesis statement, either outlining a personal belief or what the essay will focus on. The striking feature of the introductions was how similar the introductions were at Time 1 and Time 2. The students seem to have internalised a pattern that was perhaps taught to them explicitly in previous language courses.

The following excerpts, from the writing produced by Winny², are typical of the introductions found in the essays. At Time 1 (Excerpt 1), the introduction begins by stating the issue (sentence 1) and then the opposing opinions (sentences 2-3: Many people... However, some people). This is then followed by stating the focus of the essay (sentence 4), a restatement of the given topic, rather than an explicit thesis.

Excerpt 1:

Nowadays the issue about whether animals should be killed and used for animal purposes is rising in the society¹. Many people argued that animals have equal right to human, therefore they can not be used for human purposes². However, some people also argued that humans are more valuable than animals, so the equal rights between animals and humans are not exist³. This essay will discuss about whether animals should have equal rights to humans by exploring some areas⁴. [Winny, T1]

A similar pattern emerges at Time 2. Winny begins her introduction with a statement of the issue (Sentence 1), and the subsequent two sentences present the controversy (Many people... However some people). The main difference seems to be that at Time 2, the sentences are more succinct. The final sentence describes what this essay will discuss, but unlike T1, here the sentence includes some elaboration rather than a rephrasing of the given topic.

Excerpt 2:

There is an issue about animal rights rising in the society today¹. Many people believe that animals should not be killed, even for a scientific research². However, some also believe that humans have a lot of benefit from those scientific research³. This essay will discuss about several reasons on

*whether or not animals should be killed even for the scientific research*⁴.
[Winny, T2]

Unlike the introductions, there were discernable changes in the conclusions produced at Time 2 compared to Time 1. At Time 1, only about a third of the essays had a suitable conclusion (n=7). At Time 2, two thirds of the essays had suitable conclusions (n=14); that is, conclusions which related to the arguments discussed in the body of the essay.

The conclusions taken from the essays written by Nguyen illustrate this difference. Whereas at Time 1 (Excerpt 3), the conclusion seemed to restate the main arguments, at Time 2 (Excerpt 4), the conclusion is more forceful and direct.

Excerpt 3:

In conclusion, we should have balance in using animals to serve different needs of humans. It is a moral principal that we should pay attention to animals' needs.

We don't give them exactly what human have, but we should give equal consideration to their needs. [Nguyen, T1]

Excerpt 4:

In conclusion, the use of animals serving humans' benefit should be carefully considered to ensure that we still give right to them. Being part of the nature, animals should be given equal consideration. However it doesn't mean that we treat them exactly the same way as humans or grant them exactly the same rights. We just equally consider their needs. [Nguyen, T2]

Similarly, the essay produced by Dony at Time 2 also shows a better conclusion. At Time 1 (Excerpt 5), the conclusion does not relate fully to the discussion. Whereas the body of the essay discussed the various uses and abuses of animals, the conclusion focuses only on medical research and need for legislation. At Time 2 (Excerpt 6), the essay concludes with a summary of all the changes required in our attitude and treatment of animals.

Excerpt 5:

In conclusion "choose the lesser of two devils", it is impossible to stop the scientific experiments on animals, whereas, some legislation should be enforced to protect the animals rights from being mistreated by commercial purpose and entertainment. [Dony, T1]

Excerpt 6:

In conclusion, we have to stop those mistreatments of animals for merely commercial profits, respect animals rights) meanwhile we also need to understand the value of animal based work by the scientists. [Dony, T2]

The most noticeable improvement in terms of Communicative Fluency was in structuring paragraphs and coherence in the development of ideas. The excerpts below, taken from the writing by Fei, illustrate the nature of this improvement. At Time 1, Fei's arguments are difficult to follow, as two arguments are merged into the one paragraph. Thus in the first sentence of this paragraph, Fei notes the differences between animals and humans, and what the differences could mean for animal based research results. This is followed by a restatement of this impact, rather than the development of the idea. In sentence 3, discussion of the evolutionary process is not clearly linked to the main idea. The last two sentences of this paragraph describe the differences between humans and animals.

Excerpt 7:

And as we know, animals should have some difference from people, which would lead to wrong results of researches¹. We should always suspect the explanations of animal based researches². The evolution of human from monkey is a long and even complex period³. People was separated from other animals by certain reasons⁴. As we have higher intelligence and speaking abilities, we are given to the priority of living in the world⁵. [Fei, T1]

In contrast, at Time 2, the main argument is clearly stated in the opening sentence of this paragraph, the topic sentence. The subsequent sentence, although beginning with an inappropriate linking phrase, does develop this argument, by stating what distinguishes men and animals. The third sentence elaborates on these differences between humans and animals. The final two sentences conclude and rephrase the argument presented.

Excerpt 8:

The most important point is that human being is different from any other animals¹. Ignoring this gap between distinct people, they all have much higher intelligence than the most clever animal, no matter it is monkey or dog². According to this feature, human has got other special abilities, such as using languages and tools³. These abilities are the most significant differences between human and animals⁴. In another word, human has some emotions, which are the lack of animals. Emotion is also a feature of high intelligence⁵. [Fei, T2]

A frequent means of creating cohesion at both Time 1 and 2 was the use of logical linkers such as ‘therefore, moreover, on the other hand’. Milton and Tsang (1993) in Shaw and Liu (1998), who analysed a corpus of Hong Kong university students’ writing, reported an overuse of such logical linkers and attributed this to the learners’ training, and the overemphasis on the use of conjuncts in English language classes. As Table 3 shows, at Time 1, the average number of sentences was 22.15, and of those 27% (about 6 out of 22 sentences) began with logical linkers. In the case of 10 learners, over 30% of their sentences began with a logical linker. At Time 2, a similar pattern emerges, both in terms of average number of sentences (22.25) and use of logical linkers (5.55). That is, about 25% of all sentences at Time 2 began with a logical connector, although only 6 learners began over 30% of their sentences with logical linkers. Very few enumerators were used (e.g., firstly, secondly). The use of more sophisticated means of linking ideas such as the use of lexical chains (repetitions and synonyms) or parallel structures was rare in the writing at Time 1 and Time 2.

Table 2: No of sentences and use of logical linkers

	<i>Time 1</i>	<i>Time 2</i>
Average No. of sentences	22.15	22.25
Range	12 - 33	13 - 30
Average No. of sentences beginning with logical connectors	5.9	5.55

Content

The essays of 16 learners showed gain on scores of Content, with three students showing a gain of two band scores. A closer analysis of the content of all scripts revealed that although the total number of ideas was similar at Time 1 (n=61, range 2-5) and Time 2 (total 65, range 1-5), what distinguished the writing of those who showed improvement on content was in terms of the development of ideas. This perhaps explains why of the 16 learners who showed improvement on Content, 14 also showed improvement on Communicative Fluency scores.

The close link between communicative fluency and content was evident in Excerpts 7 and 8 above and in the following two excerpts. As Excerpt 9 shows, Shu presented a number of arguments as to why animals should not be given equal rights to humans: animals are useful in medical testing (sentences 2, 7, 8), humans ability to convey feelings (sentence 3), the notion of rights (sentence 4), pragmatic considerations (sentence 5), animals’ inability to speak (sentence 6). The arguments are not well

developed, and thus the impression created is that of a string of arguments which are not related.

Excerpt 9

Animal testing is still needed in medical fields as well as cosmetics¹. With animal testing, many sickness and health problems such as diabetis and asthma can be cured (Search Volume 25, no. 9 1994)². Humans can convey feelings and expression clearly showing likes and dislikes of each individual³. For example, child rights; labour rights; women rights cannot be allocated to animals⁴. To give animals equal rights as human is not only a joke but not practical as well⁵. An animal cannot speak for themselves or protest against any dislike or mistreat⁶. We can always stop animal testing in medical field, yet when in extreme need, no one protect against the use of animals for research⁷. Certainly, one suffering from asthma would not object to being cured despite the medicine given went through animal testing!⁸ [Shu, T1]

At Time 2, Shu again presents a number of arguments in support of treating animals differently, but this time each argument is elaborated and is linked to an overarching main argument (sentence 1). As Excerpt 10 shows, there is improved coherence and cohesion as the entire paragraph is devoted to supporting the main argument against the use of animals in medical research, rather than including a string of disjointed ideas as was the case at Time 1.

Excerpt 10:

Just because human are unique in the sense that they can speak and think intelligently does not justify the actions of using animals as research subjects¹. Why should animals have to endure lab testing of chemicals just for the benefit of human?² if we need to know the effectiveness of a drug or the side effects of it, we, humans, should be the subject of experiment, not the animals³. Animal's genes are after all different from human⁴. Besides, we should be prepared to undergo risk and take responsibility in testing the consequences of our own medicines⁵. Years of scientific experiment on animals could still produce long term side effect on a drug declared safe for human consumption⁶. It is ironic that human should place so much importance on themselves and declare their uniqueness, and yet still depend on animals as experiment subjects⁷. If we are so special and no animals are equal to us, then why the use of animals for testing?⁸ Humans should be the specimens instead⁹. [Shu, T2]

Language Use

The number of students whose writing showed an improvement on Form was 14, with only two showing improvement of two band scores. Form collapses accuracy and vocabulary use, and thus a closer analysis

attempted to investigate which aspects of language use showed development over time.

Fluency

The analysis based on word count revealed that the texts produced at Time 2 were on average slightly shorter: 380.60 words at Time 1 compared to 379.40 words at Time 2. Although the average number of T-units remained the same, these T-units were also on average shorter (15.09 at Time 1 compared to 14.55 at Time 2).

Table 3: Measures of fluency

	Time 1	Time 2
Average No. of words	385.60	379.40
Range:	255 - 560	237 - 491
Average No. of T-units	25.95	25.90
Range:	16 - 36	13 - 32
Average W/T	15.09	14.55

Accuracy

Accuracy scores and distribution of errors are summarised in Table 4. As the Table suggests, there was a slight improvement in accuracy scores, reflected in all measures of accuracy, but the distribution of errors remained unchanged. Most of the errors were morphological (use of verbs, marking of plural nouns, and use of articles).

Table 3: Measures of fluency

	Average No. errors	Errors/words	EFT/T (%)	EFC/C (%)	Distribution of total errors (% of total errors)		
					Morph	Syntax	Lexis
Time 1	32.70	0.08	35%	46%	61%	14%	25%
Time 2	30.50	0.06	38%	51%	61%	13%	26%

Table 4: Accuracy scores and type of errors

Complexity

The texts were also analysed for syntactic complexity. Table 5 summarises the results for the range of syntactic complexity scores used in this study, showing a slight decrease in syntactic complexity (C/T and DC/C). What these results suggest is that the slight improvement in accuracy may have been achieved due to reduced syntactic complexity and reduced length in words (see Table 3).

Table 5: Syntactic complexity

	<i>Time 1</i>	<i>Time 2</i>
C/T	1.59	1.56
DC/C	0.37	0.35

Table 6 presents a comparison of the average word length at Time 1 and Time 2. It also presents the frequencies of lexical items considered rich and those considered informal and vague. As the Table shows, word length remained fairly stable over time, as did the total and average number of lexically rich words. However, the level of formality increased. At Time 2, the learners produced fewer informal, chatty expressions, and used fewer vague words, first and second person pronouns and contractions. Announcements such as “I’m going to write about” were replaced with “This essay will discuss” and personal anecdotes introduced by phrases such as “I’ve seen lots of hosts” were less prevalent in the learners’ writing at Time 2.

Table 6: Lexical complexity (richness and informality)

	<i>Time 1</i>	<i>Time 2</i>
Average length of words (Characters/words)	4.85	4.93
Rich vocabulary		
Total No.	153	143
Average	7.65	7.15
Informal register		
Total No.	133	98
Average	6.65	4.9

DISCUSSION

A comparison of essays produced at Time 1 and Time 2 showed that the learners’ writing improved in terms of mastery of writing style. The majority of essays at Time 2 improved in terms of communicative fluency and content, particularly in presenting more well developed and cohesive arguments, and more appropriate conclusions. These improvements were reflected in the global scores as well as in the more detailed qualitative analysis.

The scores on the criterion of Form showed that a majority of students (12) showed improvement. However, this compares to 16 who showed improvement on Communicative Fluency and Content scores. Furthermore, a closer analysis of language use showed mixed results. Fluency (in terms of length) did not improve. This is perhaps not surprising

given that the participants were given instructions as to length (over 300) and a relatively short time to produce the writing. There was a slight improvement in accuracy, but this may have been because the texts produced at Time 2 were shorter and less syntactically complex. Language produced was more formal, with learners refraining from using informal expressions and first person singular pronouns, but lexical richness showed no improvement.

Increased formality in learners' writing was also reported by Shaw and Liu (1998), who note that despite the learners' greater exposure to informal spoken language when studying in an English speaking country, there was a shift to formal language expressions in the learners' writing following an EAP course. In this study, too, greater exposure to oral language whilst studying in Australia, did not lead the learners to adopt a more informal style of writing. Learners in this study were mainly postgraduate students who are required to read a large volume of texts written in a formal register, and therefore were perhaps less likely to be influenced by an informal writing style most often encountered in the media.

However, one could argue that this large amount of reading the students are required to do would have an impact on the range of vocabulary they display in their writing. Thus, the finding that lexical richness did not improve at Time 2 was somewhat surprising. One reason for the lack of evidence of improved lexical richness may be related to the topic of the essay. The topic the learners wrote about did not relate to their area of study. We are perhaps more likely to see an improvement in lexical richness when students write on topics which are related to their study major. Our current research project investigates the development of students' writing when writing on a topic related to their area of study and interest.

The lack of improvement in linguistic complexity and the only slight improvement in accuracy is of some concern. These findings show once again that mere exposure to an input-rich environment is insufficient for improving linguistic proficiency (e.g., Hinkel, 2003). As research in immersion contexts has shown (e.g., Swain, 1991) development of linguistic accuracy and complexity requires learners to be pushed to produce accurate language and to receive feedback on their language. Ferris (2003) argues that for language learners, feedback on writing may be the single most important element that affects their successful development as writers. Although university students are generally required to produce lengthy assignments for assessment, language use and accuracy are often allocated a small proportion of the overall grade (Storch & Tapper, 2000). Furthermore, in our larger project, interview data showed that for a number of these students, assessment involved limited opportunities for writing practice. Students reported being asked to produce group reports where

each individual learner's contribution was limited, or where assignment comprised calculations rather than written discourse (see Storch & Hill, under review).

CONCLUSION

What this study has shown is that, for the majority of these students, studying in an English medium university and being immersed in the L2 did lead to language improvement, even after just one semester. However, the improvement was evident mainly in the macro level features: improved cohesion, more appropriate conclusions, and better developed arguments. Language use, particularly in terms of linguistic accuracy and sophistication showed no or little improvement.

The results of this study need to be interpreted cautiously. The study was small scale and this precluded the use of robust statistical tests of significance. Furthermore, the averages presented for the cohort disguise the fact that some students improve more than others. It is also important to note that one of the problems with studies using a pre-post test format is that they are based on the assumption that all participants will be equally motivated to complete the test to the best of their ability on both occasions. Test takers tend to perform better on a test when the results have high stakes (e.g., lead to important decisions) compared to low stakes (e.g., practice tests).

Nevertheless, the findings of this study suggest that mere immersion in an English medium university may not lead to development in linguistic proficiency of international students, even after one semester of study. The findings suggest that learners may need to be encouraged to attend ESL writing classes, where the focus is on improving students' academic writing skills, particularly in terms of linguistic accuracy and sophistication.

NOTES :

¹ Another commonly used measure of lexical complexity is a type-token ratio (Ellis & Barkuizen, 2005). This measure considers the total number of different words used (and thus excludes all repeated words) as a proportion of all words in the text produced. However, the measure is best used with texts of over 350 words. In this study, many of the texts were below 350 words in length.

² All names used are pseudonyms.

REFERENCES

- Bardovi-Harlig, K. & Bofman, T. 1989. Attainment of syntactic and morphological accuracy by advanced language learners. *Studies in Second Language Acquisition*, 11, 17-34.
- Biber, D. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Cumming, A., Kantor, R., Baba, K., Erdosy, U., Eouanzoui, K. & James, M. 2005. Differences in writing discourse in independent and integrated prototype tasks for next generation TOEFL. *Assessing Writing*, 10, 5-43.
- Elder, C. & O'Loughlin, K. 2003. Investigating the relationship between intensive EAP training and band score gains on IELTS. *IELTS Research Reports*, 4 (pp. 207-254). Canberra: IELTS Australia.
- Ellis, R. & Barkhuizen, G. 2005. *Analysing learner language*. Oxford: Oxford University Press.
- Ferris, D. 2003. *Response to student writing: Implications for second language students*. Mahwah, N. J.: Lawrence Erlbaum Associates.
- Foster, P. & Skehan, P. 1999. The influence of sources of planning and focus of planning on task-based performance. *Language Teaching Research*, 3/3, 215-247.
- Green, A. & Weir, C. 2003. Monitoring score gain on the IELTS Academic Writing module in EAP Programmes of varying duration. Phase 2 report, Cambridge: UCLES.
- Hinkel, E. 2003. Simplicity without elegance: Features of sentences in L1 and L2 academic texts. *TESOL Quarterly*, 37/2, 275-301.
- Hunt, K. 1966. Recent measures in syntactic development. *Elementary English*, 43, 732-139.
- Krashen, S. D. 1985. *The input hypothesis: Issues and implications*. London: Longman.
- Polio, C. 1997. Measures of linguistic accuracy in second language writing research. *Language Learning*, 47/1, 101-143.
- Read, J. & Hays, B. 2003. The impact of the IELTS test on preparation for academic study in New Zealand. *IELTS Research Reports*, 5 (pp. 237-262). Canberra: IELTS Australia.

- Richards, J. C. Platt, J., & Platt, H. 1992. *Dictionary of language teaching and applied linguistics*. Essex: Longman
- Shaw, P. & Liu, E. T-K. 1998. What develops in the development of second language writing. *Applied Linguistics*, 19/2, 225-254.
- Storch, N. & Hill, K. (under review). What happens to international students' English after one semester at university?
- Storch, N. & Tapper, J. 2000. Discipline specific academic writing: What content teachers comment on. *Higher Education Research & Development*, 19 /3, 337-356.
- Swain, M. 1985. Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. M. Gass & C. G. Madden (Eds.), *Input in second language acquisition* (pp. 235-256). Rowley, MA: Newbury House.
- Swain, M. 1991. Manipulating and complementing content teaching to maximise second language learning. In R. Phillipson, E. Kellerman, L. Selinker, M. Sharwood-Smith, & M. Swain (Eds). *Foreign/second language pedagogy research. A commemorative volume for Claus Faerch*. Clevedon: Multilingual Matters.
- Swales, J. 1985. *Episodes in ESP*. Oxford: Pergamon.
- Wolf-Quintero, K., Inagaki, S. & Kim, H. 1998. *Second language development in writing: Measures of fluency, accuracy and complexity*. Honolulu, Hawaii: University of Hawaii at Manoa.